ED 402 340 TM 025 927

AUTHOR	Sheehan, Janet K.; Han, Tianqi
TITLE	How Do Extreme Schools Change the Interpretation of Results in School Effectiveness Research? Effects of Outlying Second-Level Variables in HLM.
PUB DATE	Apr 96
NOTE	12p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Effective Schools Research; Elementary Secondary Education; *Error of Measurement; *Estimation (Mathematics); *School Effectiveness; Simulation; *Statistical Analysis
IDENTIFIERS	Fixed Effects; *Hierarchical Linear Modeling; Multilevel Analysis; *Outliers

ABSTRACT

This study examined the robustness of the estimation of fixed-effects in multilevel analysis, as might occur in the conduct of school-effects studies with outlying schools. Outlying values for both intercepts and slopes for individual schools were modeled separately to determine the effects of extreme values of second-level variables on the fixed-effect parameter estimation. A total of seven data sets were generated for the simulation. Under the conditions investigated in this study, adding a cluster of outlying schools had little effect on the estimation of gamma 10 and gamma 11. However, the standard error of gamma 01 increased, thereby increasing the conservativeness of the test of significance of gamma 01. This occurred for clusters of outlying slopes, intercepts, or a combination of both. Introducing a single outlying school also increased the standard error of gamma 01, the effects being more dramatic when the outlier was an extreme slope. (Author/SLD)

How do Extreme Schools Change the Interpretation of Results in School-Effectiveness Research?: Effects of Outlying Second-Level Variables in HLM.

Janet K. Sheehan & Tianqi Han
Northern Illinois University

This paper was presented at the Annual Meeting of the American Educational Research Association, New York, NY, April, 1996

Abstract

This study examined the robustness of the estimation of fixed-effects in multilevel analysis, as might occur when conducting school-effects studies with outlying schools. Outlying values for both intercepts and slopes for individual schools were modeled separately to determine the effects of extreme values of second-level variables on the fixed-effect parameter estimation. Under the conditions investigated in this study, adding a cluster of outlying schools had little effect on the estimation of γ_{10} and γ_{11}, however the standard error of γ_{01} increased, thereby increasing the conservativeness of the test of significance of γ_{01}. This occurred for clusters of outlying slopes, intercepts, or a combination of both. Introducing a single outlying school also increased the standard error of γ_{01}, the effects being more dramatic when the outlier was an extreme slope.

How do Extreme Schools Change the Interpretation of Results in School-Effectiveness Research?: Effects of Outlying Second-Level Variables in HLM.

Theoretical Perspective

Multilevel analysis has recently replaced multiple linear regression,(MLR) as the method of choice for school-effects research (Mendro, Webster, Bembry, & Orsak, 1995; Webster, Mendro, & Almaguer, 1993). Bryk and Raudenbush, (1992) give at least two explanations for the appropriateness of multilevel analysis for school-effects research. First, random variation and structural effects may exist at more than one level, and therefore a correctly specified model is a multilevel model in which fixed and random effects can be estimated at each level. Second, the assumption of independence of errors in MLR is violated when there is intraclass correlation, as one would find in such hierarchically nested data.

The models for a simple multilevel analysis with one level-1 and one level-2 fixed-effect variable would be:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1j} + r_{ij} \quad (1)$$

for the level-1 model and

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j} \quad (3)$$

for the level-2 models. The level-1 model is analogous to a simple regression model, where r_{ij} is the level-1 error. The parameters from level-1 become outcome variables in the level-2 models, which are predicted by second-level effects. u_{0j} and u_{1j} are the random error components of level 2. Upon inspection of these models, it is apparent that aberrant or extreme values for β_{0j} and β_{1j} will affect the estimation of parameters in the second-level models. These discrepant schools could have an undue influence on the estimation of the second-level fixed effects of the model. Therefore, Bryk and Raudenbush, (1992) suggested that cross-level exploratory analyses be performed when conducting school-effects studies to determine if there are schools with outlying intercepts or slopes.

The effects of outliers have been extensively studied in ordinary least squares regression analysis. Rousseeuw and van Zomeren (1990) identify three types of outliers: a) vertical outliers in which there is a large residual, b) leverage point outliers, which are consistent with the relationships found in the rest of the data and c) leverage point outliers, which significantly alter the relationships by their inclusion, as well as having high leverage. Outliers exert effects in not only the results of the main and interaction-effects significance tests from GLM, but also in the estimation of the model parameters (Douzenis & Rakow, 1987; Hecht, 1991).

This study is designed to examine the robustness of the estimation of fixed-effects in multilevel analysis, when conducting school-effects studies with outlying schools. Outlying values for both intercepts and slopes for individual schools are modeled separately to determine the effects of extreme values of second-level variables on the fixed-effect parameter estimation.

Data Source

A total of seven data sets were generated for this study. Data set I was simulated to serve as the major body of data. This dataset consisted of 250 schools with 50 students in each school. Three datasets were generated based on the original dataset, but containing 10%: a) schools with extreme intercepts, b) schools with extreme slopes, or c) schools with extreme intercepts and slopes. The extreme intercept is analogous to adding vertical outliers. These were generated by increasing the mean approximately 3 standard deviations units higher. The slope outliers were generated by reducing the correlation between X and Y from .7 to .4, holding s_x and s_y constant. The dataset size was maintained at 250 level II observations or schools. The three new datasets were termed Outlier-I, Outlier-S, and Outlier-SI, respectively. Multiple regression diagnostics tests of the schools, were examined to determine the most extreme outliers for each of the three datasets, as well as to determine that each dataset contained 10% outliers. Specifically, the intercept and slope DF Betas were checked to ensure that they increased for the outlying observations.

To produce the last three data sets, extreme outlying observations were substituted for observations in the original dataset. One school was removed from the original data set, and the most extreme outlying school found in the analysis of Outlier-I was inserted in its place. This new dataset was termed Single-I. This process was repeated for Outlier-S, and Outlier-SI, producing the new datasets Single-S, and Single-SI, respectively.

Procedures and Results

A two-level hierarchical linear modeling program (HLM) was used to conduct the multilevel analysis for this study. Separate HLM analyses were conducted for each dataset. Table 1 presents the descriptive statistics of the seven datasets. It can be seen that adding the 10% outlying schools increased the standard deviations of the original dataset. However, adding the one outlying school had little effect on the standard deviations.

Table 1

Descriptive Statistics

	Outlier	Variable	N	Mean	SD	Minimum	Maximum
No Outlier		X	12,500	180.04	9.04	142.51	214.02
		Y	12,500	192.03	13.02	140.30	241.76
		W	250	5.60	0.14	5.21	5.99
10% Outliers	Intercept	X	12,500	180.04	9.04	142.51	214.02
		Y	12,500	197.03	19.34	140.30	282.38
		W	250	5.60	0.14	5.21	5.99
	Slope	X	12,500	180.02	9.06	142.51	214.02
		Y	12,500	184.20	26.89	73.44	241.76
		W	250	5.60	0.14	5.21	5.99
	Intercept & Slope	X	12,500	180.05	9.07	142.51	214.02
		Y	12,500	188.27	17.25	106.95	241.76
		W	250	5.60	0.14	5.21	5.99
Single Outlier	Intercept	X	12,500	180.04	9.04	142.51	214.02
		Y	12,500	192.23	13.33	140.30	276.10
		W	250	5.60	0.14	5.21	5.99
	Slope	X	12,500	180.05	9.05	142.51	214.02
		Y	12,500	191.73	13.87	94.20	241.76
		W	250	5.60	0.14	5.21	5.99
	Intercept & Slope	X	12,500	180.05	9.04	142.51	214.02
		Y	12,500	191.89	13.22	127.53	241.76
		W	250	5.60	0.14	5.21	5.99

Table 2

Average Regression Coefficient and Reliability Estimation for Level I

	Outlier Type	β_0	Reliability$_{\beta 0}$	β_1	Reliability$_{\beta 1}$
No. Outlier		192.00	0.486	1.01	0.043
	Intercept	197.00	0.992	1.01	0.046
10%	Slope	184.20	0.997	0.97	0.465
Outliers	Intercept & Slope	188.30	0.986	0.97	0.432
	Intercept	192.20	0.853	1.01	0.053
Single	Slope	191.70	0.936	1.01	0.059
Outlier	Intercept & Slope	191.90	0.803	1.01	0.080

Table 2 presents the level I regression coefficients and the corresponding reliability estimates.[1] It can be seen that the reliability estimates for all the datasets with outliers were increased. The increase in the school differences in the parameter estimates increased ι_{qq} which increased the reliability estimates. The average parameter estimate for β_0 changed when outliers were added. This occurred for both 10% outliers and single outliers. The average coefficient estimate for β_1 did not change when intercept outliers were added. However, it did change the average estimate of β_1 when 10% slope or slope and intercept outliers were added.

Table 3

Regression Coefficient for Level II

	Outlier Type	γ_{00}	SE γ_{00}	γ_{01}	SE γ_{01}	γ_{10}	SE γ_{10}	γ_{11}	SE γ_{11}
No Outlier		192.0*	0.116	4.51*	0.804	1.01*	0.009	0.07	0.066
	Intercept	197.0*	0.911	-10.7	6.333	1.01*	0.010	0.07	0.066
10%	Slope	184.2*	1.498	1.06	10.412	0.97*	0.013	0.05	0.091
Outliers	Intercept & Slope	188.3*	0.726	2.54	5.048	0.97*	0.013	0.08	0.089
	Intercept	192.2*	0.217	4.18*	1.505	1.01*	0.010	0.06	0.067
Single	Slope	191.7*	0.328	2.15	2.277	1.01*	0.010	0.06	0.067
Outliers	Intercept & Slope	191.9*	0.187	4.21*	1.301	1.01*	0.010	0.07	0.068

Table 3 shows the regression coefficients estimates for level II, equations 2 and 3. The standard error estimates of all parameters increased relative to the dataset with no outliers. The outliers had little effect on both γ_{11} and γ_{10} and did not alter the statistical test results. This indicates that the outliers generated under these conditions had little effect on the relationship between X and Y when W=0, γ_{10}, or the mean difference in the X-Y slopes across levels of W, γ_{11}. However, the outliers did have an effect on γ_{01}, the mean difference in Y across levels of W. Both the values for γ_{01} and its standard error changed such that in all datasets with 10% outliers the statistical test of γ_{01} was no longer significant. The net effect of the outliers then, was making the tests of γ_{01} more conservative. For the single outlier datasets, only the outlying slope changed the results of the significance tests of γ_{01}.

Discussion

When a cluster of outlying level-II observations are added to a multilevel analysis, the effects on the estimation of level II fixed effects can be dramatic. In school-effectiveness research, this would be manifested if a cluster of schools had extreme means on Y, aberrant relationships between X and Y in a cluster of schools, or if a cluster of schools had both extreme means and extreme X-Y relationships. Under the conditions investigated in this study the changes in the parameter estimates and their standard errors resulted in conservative tests of significance. Adding a single outlying school naturally had less of an effect, however, results of tests of significance were still changed in the presence of a single outlying slope. Further, it is conceivable that similar results could arise for single outliers of the intercept and the slope and intercept for more extreme outliers.

Future Research

This study suggests a more extensive study to ascertain how outliers of differing magnitudes affect parameter estimation in multilevel analysis is needed. It would also be of interest to study the effects of other types of outliers, such as leverage point outliers, on parameter estimation in multilevel analysis.

References

Bryk, A. S. & Raudenbush, S. W. (1992). Hierarchical Linear Models: Applications, and Data Analysis Methods. Newbury Park, CA : Sage Publications.

Douzenis, C. & Rakow, E. A. (1987, November). Outliers: A potential data problem. ERIC ED291798.

Hecht, J. B. (1991, April). Least-square linear regression and schrodinger' Cat: Perspectives on the analysis of regression residuals. ERIC ED347194.

Mendro, R. L, Webster, W. J., Bembry, L. K., & Orsak, T. H. (1995, April). An application of hierarchical linear modeling in determining school effectiveness. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Rousseeuw, P. J. & Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85, 633-639.

Webster, W. J., Mendro, R. L., & Almaguer, T. O. (1993). Effectiveness indices: The major component of an equitable accountability system, ERIC TM 019913.

Footnotes

[1]Reliability in HLM refers to the ratio of parameter variance to total variance.

$$\text{Reliability } (\hat{\beta}_{\text{q}} = \frac{1}{J}\sum_{j=1}^{J} \tau_{qq} / (\tau_{qq} + \nu_{qq}) \qquad (4)$$

for each q=0,...,Q (Bryk & Raudenbush, 1992, p.43) and where ι_{qq} = parameter variance and ν_{qq} = error variance (Bryk & Raudenbush, p.34).

AERA April 8-12, 1996

TM025927

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: How do Extreme Schools Change the Interpretation of Results in School-Effectiveness Research?: Effects of Outlying Second-Level Variables in HLM.

Author(s): Janet K. Sheehan and Tianqi Han

Corporate Source:	Publication Date:
Northern Illinois University	May 30, 1996

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document		Sample sticker to be affixed to document [ ]
Check here Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction	"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY _Sample_ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)." **Level 1**	"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY _Sample_ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)." **Level 2**
		or here Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Janet K. Sheehan	Position: Assistant Professor
Printed Name: Janet K. Sheehan	Organization: Northern Illinois University
Address: EPCSE Northern Illinois University Dekalb, IL 60115	Telephone Number: (815) 753-8523
	Date: 4/15/96

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

February 27, 1996

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a written copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (23)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: AERA 1996/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://tikkun.ed.asu.edu/aera/). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an AERA chair or discussant, please save this form for future use.